MITRE Analytic Roundtable Use Case

# Use case: Threat Analysis

## User Need:

While threat assessments are a proven approach to lower the risk of targeted violence by proactively identifying, assessing and mitigating threats, the resource intensive methods utilized can make the process prohibitive in both time and money.

Sponsors seek analytic capabilities that reduce the costs of doing complex data exploration and increase the effectiveness in identifying trends and patterns of targeted violent events over time and space. These capabilities can help sponsors’ analysts more effectively identify and assess threats in immediate time to control and neutralize specific threat actors.

## Centrifuge Systems, ESRI, Recorded Future and SAP NS2 are partnering to demonstrate integrated an analytic system that efficiently applies commercially available solutions demonstrating a threat assessment approach to counter the terrorism posed by the militant Islam Group, Boko Haram in the Sub-Saharan African region.

## Use case goal:

To assess the growing threat of Boku Haram in Africa based on data that indicates increasing attacks and increasing numbers of fatalities associated with these attacks. From 2014 to 2015, Boku Haram attacks increased along with numbers of fatalities. The ACLED data provides this information. Recorded Future information provides other insights into activity.

# Partners:

SAP

ESRI

Centrifuge Systems

Recorded Future

# Data set:

ACLED

<http://www.acleddata.com/>

## Data Description:

ACLED (Armed Conflict Location and Event Data Project) is designed for disaggregated conflict analysis and crisis mapping. This dataset codes the dates and locations of all reported political violence and protest events in over 60 developing countries in Africa and Asia. Political violence and protest includes events that occur within civil wars and periods of instability, public protest and regime breakdown. The project covers all African countries from 1997 to the present, and South and South-East Asia in real-time.

ACLED is directed by Prof. Clionadh Raleigh (University of Sussex). It is operated by senior research manager Andrea Carboni (University of Sussex) for Africa and Hillary Tanoff for South and South-East Asia. The data collection involves several research analysts, including Charles Vannice, James Moody, Daniel Wigmore-Shepherd, Andrea Carboni, Matt Batten-Carew, Margaux Pinaud, Roudabeh Kishi, Helen Morris, Braden Fuller, Daniel Moody and others.

These data contain information on:

* Dates and locations of conflict events;
* Specific types of events including battles, civilian killings, riots, protests and recruitment activities;
* Events by a range of actors, including rebels, governments, militias, armed groups, protesters and civilians;
* Changes in territorial control; and
* Reported fatalities.

Event data are derived from a variety of sources including reports from developing countries and local media, humanitarian agencies, and research publications. Please review the [codebook](http://www.acleddata.com/wp-content/uploads/2017/01/ACLED_Codebook_2017.pdf) and [user guide](http://www.acleddata.com/wp-content/uploads/2017/01/ACLED_User-Guide_2017.pdf) for additional information: the codebook is for coders and users of ACLED, whereas the brief guide for users reviews important information for downloading, reviewing and using ACLED data. A specific [user guide for development and humanitarian practitioners](http://www.acleddata.com/wp-content/uploads/2017/01/ACLED_User-Guide-for-Humanitarians_2017.pdf) is also available, as is a guide to our sourcing materials.

Data are available for public download from the data page. Data for all African countries and select South and South East Asian states are from January 1997 and are updated in real-time. Real-time monthly ACLED data for South and Southeast Asian and African countries is available for download on this or the [Climate Change and African Political Stability Project (CCAPS) website](http://www.strausscenter.org/acled.html). For more information, please contact Prof. Raleigh or the project administrator using the contact form on this website. Other data coded in the initial stage of this project are available from the data page, under various version headings.

Maps and trend analyses of ACLED data are available in several locations and forms: real-time analysis of conflicts can be found on the [ACLED Crisis Blog](http://www.crisis.acleddata.com/); trend reports which review monthly changes and patterns of political violence and protest and are available on the [Publications page](http://www.acleddata.com/research-and-publications/); maps and graphics of real-time violence are available on the [Visuals page](http://www.acleddata.com/visuals/); and in-depth research papers can be found on the [Research page.](http://www.acleddata.com/geopv/)

If using ACLED data, please cite: Raleigh, Clionadh, Andrew Linke, Håvard Hegre and Joakim Karlsen. 2010. Introducing ACLED-Armed Conflict Location and Event Data. Journal of Peace Research 47(5) 651-660.

Any questions or comments can be forwarded using the Contact Form on this website.

Annual Reports and Status

ACLED is registered as an independent, non-governmental organization in the United States. Annual reports and recent financial statements are available [here](http://www.acleddata.com/wp-content/uploads/2016/05/ACLED-Registration-and-501-c3-Documentation.zip).

# System Architecture

Our multi-vendor system uses integration to correlate and analyze multiple data types including geospatial, temporal, unstructured, and structured data, combined with utilizing visualization to understand the relationship between large quantities of events in near real-time.

ACLED data is used to link conflict actors to other conflict actors and the resultant number of fatalities within a specific country and within a defined timeframe. The data is stored in SAP HANA and visually analyzed in Centrifuge Analytics with Drill Charts, Timeline, Link Analysis, and Geospatial visualizations. Within Centrifuge Analytics, ESRI provides the map service and Geospatial Data Services for the Geospatial visualization of the conflicts based on the coordinates in the ACLED data. ESRI also provides visualizations in ESRI client applications.

Data enhancement is added by querying Recorded Future data around the dates of the event data contained in the ACLED data. The data is run through SAP’s NLP engine with sentiment analysis. This data is then linked to the ACLED data that is visualized.

The visual data analysis presents event actors, dates, event types, country, and number of fatalities. With the enhanced data, the analyst is also able to see the opinions of the events, what people in the area fear, and, possibly, any Twitter intelligence leading up to the event.

The data is used to analyze two threat types: internal and external, based on the actors and their affiliation. Centrifuge Analytics is used to visually analyze the actors and affiliations to determine the type of threat. Esri is used to visualize hot spots of activity geospatially.

# System Demonstration

The system demonstrates how an analyst can perform temporal, geospatial, structure data and unstructured data analysis more efficiently using integrated commercial solutions.

As an example, the analyst is interested in looking at concentration of events that happened in all of Africa in the last two years. He/she zooms in on the timeline until only the last 2 years’ worth of data are visible. This automatically updates the rest of the interface to filter out any events that did not occur within the selected time frame.

Now that the analysts have easily narrowed the time frame of events in view, they can go to the next step and focus on a specific geographical area. The analysts next filter geospatially by simply zooming in to the region around Nigeria, effectively filtering out any events that did not occur in that region.

Since the timeframe and region have been defined by the analysts they can now start to gain an understanding of what has actually happened in that region during that time period by looking at the tag cloud. The analysts can select a term in the tag cloud to show only events related to their selected term or they can enter a term of their own.

Recorded future analyzes and provides additional data not contained in the ACLED data set that identifies events leading up to a major event in 2015 referred to as the Fotokol Event which is a major Boku Haram attack on the town of Fotokol in Cameroon. The events signify an escalation of activity just prior to the event and battle damage assessment after the event. This data can be added to the ACLED data to give more insight on predicting the events.

With a subject matter selected the analysts can now switch the heat map to show the number of fatalities as opposed to simply the number of events within their selected region and time frame. From there they can zoom in on the map and select any event to view the structured data points that describe the event.

# Data Exchange Model

SAP ingests the ALED data along with enrichment from Recorded Future. The data is structured in SAP and provided to the Transfer service of the Data Exchange. ESRI also provides Geospatial data to the Exchange. A tool, such as Centrifuge Analytics will then extract and analyze the information from the Data Exchange.

# System Flow

         ACLED data set is ingested into **SAP/NS2 HANA**.

         HANA provides analysis of ACLED’s unstructured textual content (which some of the ACLED data entries contain) to produce sentiment analysis and entity extraction; this is performed within HANA and results will be put back into HANA.

         Queries are performed to **Recorded Future** given information identified in the structured and unstructured text of ACLED, retrieving further content that contains both unstructured and structured text to add more instances.

         ACLED also provides geo-location information, which is sent to **Esri** for further enrichment.

         These different sources, all currently in HANA, are adapted into the Exchange’s data model and moved into the Analysis Exchange (this is MITRE’s principal responsibility).

         Once in the Analysis Exchange, **Centrifuge** and **Esri** access this to visualize the linked data information and map information.

* Centrifuge expects to receive, at minimum, the following fields in JSON format from the exchange:
  + EVENT\_ID\_CNTY
  + ACTOR1
  + ACTOR2
  + ALLY\_ACTOR\_1
  + ALLYACTOR\_2
  + EVENT\_DATE
  + YEAR
  + EVENT\_TYPE
  + COUNTRY
  + LOATION
  + LATITUDE
  + LONGITUDE
  + SOURCE
  + FATALITIES

Note, in the pipeline above, several steps of enrichment are performed before any content reaches the Analysis Exchange (e.g., Recorded Future to HANA, HANA’s own text analysis). We receive the enriched results from HANA, which is what is adapted into the ontology. This is why we’ve limited these questions to just the connection points to the Analysis Exchange. There are still topics within the overall use-case that the team should address, such as what particular analytics are chosen, the orchestration between the analytics, and how the results will be represented in HANA.

# Data Flow

